



wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 16, 2024 – 08:44 PM EDT

PDB ID : 5GHS
Title : DNA replication protein
Authors : Oyama, T.
Deposited on : 2016-06-20
Resolution : 2.59 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

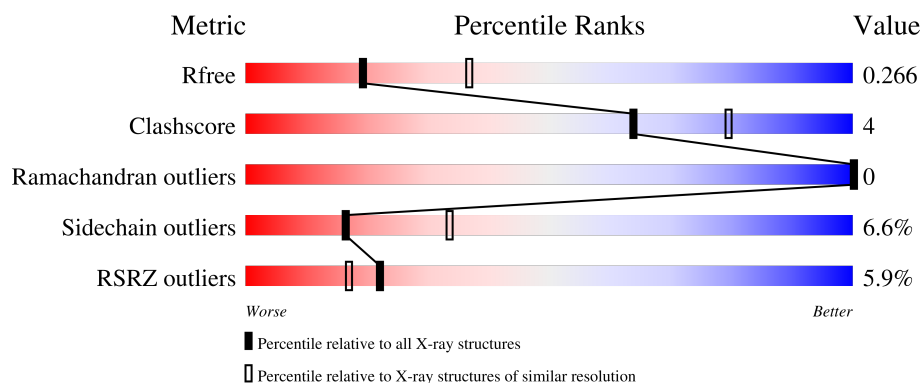
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.59 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions $\leq 5\%$. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	477	<div> <div>9%</div> <div>80%</div> <div>13%</div> <div>6%</div> </div>
1	B	477	<div> <div>2%</div> <div>62%</div> <div>9%</div> <div>28%</div> </div>
2	C	80	<div> <div>60%</div> <div>10%</div> <div>30%</div> </div>
2	D	80	<div> <div>4%</div> <div>76%</div> <div>12%</div> <div>11%</div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 7252 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called SsDNA-specific exonuclease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	450	Total	C	N	O	S	0	0	0
			3498	2207	612	667	12			
1	B	345	Total	C	N	O	S	0	0	0
			2708	1708	477	513	10			

- Molecule 2 is a protein called Putative uncharacterized protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	56	Total	C	N	O	S	0	0	0
			437	284	74	76	3			
2	D	71	Total	C	N	O	S	0	0	0
			540	344	95	97	4			

There are 44 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	109	MET	-	expression tag	UNP Q5JF31
C	110	GLY	-	expression tag	UNP Q5JF31
C	111	SER	-	expression tag	UNP Q5JF31
C	112	SER	-	expression tag	UNP Q5JF31
C	113	HIS	-	expression tag	UNP Q5JF31
C	114	HIS	-	expression tag	UNP Q5JF31
C	115	HIS	-	expression tag	UNP Q5JF31
C	116	HIS	-	expression tag	UNP Q5JF31
C	117	HIS	-	expression tag	UNP Q5JF31
C	118	HIS	-	expression tag	UNP Q5JF31
C	119	SER	-	expression tag	UNP Q5JF31
C	120	SER	-	expression tag	UNP Q5JF31
C	121	GLY	-	expression tag	UNP Q5JF31
C	122	GLU	-	expression tag	UNP Q5JF31
C	123	ASN	-	expression tag	UNP Q5JF31
C	124	LEU	-	expression tag	UNP Q5JF31

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	125	TYR	-	expression tag	UNP Q5JF31
C	126	PHE	-	expression tag	UNP Q5JF31
C	127	GLN	-	expression tag	UNP Q5JF31
C	128	GLY	-	expression tag	UNP Q5JF31
C	129	HIS	-	expression tag	UNP Q5JF31
C	130	MET	-	expression tag	UNP Q5JF31
D	109	MET	-	expression tag	UNP Q5JF31
D	110	GLY	-	expression tag	UNP Q5JF31
D	111	SER	-	expression tag	UNP Q5JF31
D	112	SER	-	expression tag	UNP Q5JF31
D	113	HIS	-	expression tag	UNP Q5JF31
D	114	HIS	-	expression tag	UNP Q5JF31
D	115	HIS	-	expression tag	UNP Q5JF31
D	116	HIS	-	expression tag	UNP Q5JF31
D	117	HIS	-	expression tag	UNP Q5JF31
D	118	HIS	-	expression tag	UNP Q5JF31
D	119	SER	-	expression tag	UNP Q5JF31
D	120	SER	-	expression tag	UNP Q5JF31
D	121	GLY	-	expression tag	UNP Q5JF31
D	122	GLU	-	expression tag	UNP Q5JF31
D	123	ASN	-	expression tag	UNP Q5JF31
D	124	LEU	-	expression tag	UNP Q5JF31
D	125	TYR	-	expression tag	UNP Q5JF31
D	126	PHE	-	expression tag	UNP Q5JF31
D	127	GLN	-	expression tag	UNP Q5JF31
D	128	GLY	-	expression tag	UNP Q5JF31
D	129	HIS	-	expression tag	UNP Q5JF31
D	130	MET	-	expression tag	UNP Q5JF31

- Molecule 3 is SULFATE ION (three-letter code: SO4) (formula: O₄S).

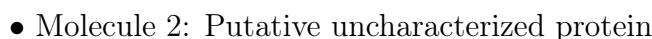


Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		
3	A	1	Total	O	S	0	0
			5	4	1		
3	B	1	Total	O	S	0	0
			5	4	1		
3	D	1	Total	O	S	0	0
			5	4	1		

- Molecule 4 is water.

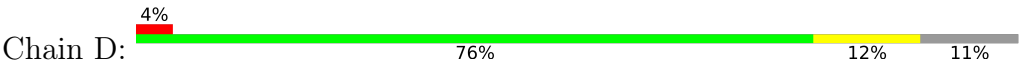
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	24	Total	O	0	0
			24	24		
4	B	14	Total	O	0	0
			14	14		
4	C	2	Total	O	0	0
			2	2		
4	D	4	Total	O	0	0
			4	4		

- Molecule 1: SsDNA-specific exonuclease





● Molecule 2: Putative uncharacterized protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	48.04Å 116.28Å 235.40Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	44.48 – 2.59 44.48 – 2.59	Depositor EDS
% Data completeness (in resolution range)	99.6 (44.48-2.59) 99.7 (44.48-2.59)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	5.19 (at 2.58Å)	Xtriage
Refinement program	PHENIX 1.7.3 _928	Depositor
R, R_{free}	0.237 , 0.273 0.230 , 0.266	Depositor DCC
R_{free} test set	2126 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å ²)	32.9	Xtriage
Anisotropy	0.745	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 38.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	7252	wwPDB-VP
Average B, all atoms (Å ²)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 8.58% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z > 5$	RMSZ	$\# Z > 5$
1	A	0.22	0/3554	0.40	0/4793
1	B	0.22	0/2754	0.40	0/3712
2	C	0.23	0/445	0.41	0/602
2	D	0.22	0/551	0.42	0/745
All	All	0.22	0/7304	0.40	0/9852

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3498	0	3452	33	0
1	B	2708	0	2697	18	0
2	C	437	0	458	4	0
2	D	540	0	533	5	0
3	A	15	0	0	0	0
3	B	5	0	0	0	0
3	D	5	0	0	0	0
4	A	24	0	0	1	0
4	B	14	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	C	2	0	0	0	0
4	D	4	0	0	0	0
All	All	7252	0	7140	60	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 4.

The worst 5 of 60 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:425:ALA:HB1	1:A:463:ALA:HB1	1.67	0.75
1:B:263:MET:HE2	1:B:276:LEU:HD22	1.75	0.68
1:A:196:GLU:OE1	1:A:202:ARG:NH1	2.33	0.61
1:A:202:ARG:HD3	1:A:294:GLU:HG3	1.85	0.58
1:B:25:HIS:CE1	1:B:77:GLU:HG2	2.40	0.56

There are no symmetry-related clashes.

5.3 Torsion angles [i](#)

5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	444/477 (93%)	427 (96%)	17 (4%)	0	100	100
1	B	343/477 (72%)	332 (97%)	11 (3%)	0	100	100
2	C	54/80 (68%)	54 (100%)	0	0	100	100
2	D	69/80 (86%)	64 (93%)	5 (7%)	0	100	100
All	All	910/1114 (82%)	877 (96%)	33 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	357/383 (93%)	334 (94%)	23 (6%)	17	35
1	B	281/383 (73%)	259 (92%)	22 (8%)	12	25
2	C	45/66 (68%)	43 (96%)	2 (4%)	28	53
2	D	54/66 (82%)	52 (96%)	2 (4%)	34	60
All	All	737/898 (82%)	688 (93%)	49 (7%)	16	33

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	91	LEU
1	B	248	ARG
1	B	140	VAL
1	B	202	ARG
1	B	253	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates [i](#)

There are no monosaccharides in this entry.

5.6 Ligand geometry [i](#)

5 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
3	SO4	D	201	-	4,4,4	0.14	0	6,6,6	0.14	0
3	SO4	A	501	-	4,4,4	0.13	0	6,6,6	0.08	0
3	SO4	A	502	-	4,4,4	0.14	0	6,6,6	0.12	0
3	SO4	A	503	-	4,4,4	0.14	0	6,6,6	0.09	0
3	SO4	B	501	-	4,4,4	0.13	0	6,6,6	0.11	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers [i](#)

There are no such residues in this entry.

5.8 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Fit of model and data [i](#)

6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å ²)	Q<0.9
1	A	450/477 (94%)	0.46	43 (9%) 8 5	11, 32, 77, 112	0
1	B	345/477 (72%)	0.23	8 (2%) 60 54	13, 28, 58, 85	0
2	C	56/80 (70%)	0.15	0 100 100	16, 30, 49, 73	0
2	D	71/80 (88%)	0.35	3 (4%) 36 29	17, 37, 72, 83	0
All	All	922/1114 (82%)	0.35	54 (5%) 22 17	11, 30, 72, 112	0

The worst 5 of 54 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	419	GLY	10.1
1	A	411	THR	7.1
1	A	347	ILE	6.7
1	A	417	GLU	6.5
1	B	343	ARG	5.8

6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates [i](#)

There are no monosaccharides in this entry.

6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled ‘Q< 0.9’ lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
3	SO4	A	502	5/5	0.91	0.16	63,65,73,90	0
3	SO4	A	503	5/5	0.91	0.34	59,66,69,89	0
3	SO4	A	501	5/5	0.94	0.18	37,42,49,53	0
3	SO4	D	201	5/5	0.95	0.14	54,65,72,77	0
3	SO4	B	501	5/5	0.96	0.17	32,38,48,52	0

6.5 Other polymers [i](#)

There are no such residues in this entry.